A WWW-based Archive and Retrieval System for Multimedia Production

Jason Hyon, Sugi Sorensen, Mike Martin, Kristy Kawasaki

Jet Propulsion Laboratory, Science Data Management and Archiving Section
California Institute of Technology
4800 Oak Grove Dr., Pasadena, CA. 91109

Mark Takacs

Netscape Communications Corporation
685-2 E Middlefield Rd., Mountain View, CA. 94043

ABSTRACT

This paper describes the Data Distribution Laboratory (DDL) and discusses issues involved in building multimedia CD-ROMs. It describes the modeling philosophy for cataloging multimedia products and the worldwide-web (WWW) based multimedia archive and retrieval system (Webcat) built on that model.

Keywords: metadata formats multimedia WWW jukebox order CD ROM

1.0 INTRODUCTION

The Data Distribution Laboratory (DDL) is a NASA facility chartered to evaluate and apply new information storage and distribution technology to real user problems. Over the past few years the DDL has produced multimedia CD ROMs such as "Welcome to the Planets", "The IGARSS Conference Proceedings CD-ROM", "The TOPEX/Poseidon Informational CD-ROM", and the "NASA Scatterometer (NSCAT) Educational CD ROM". The lab is currently working on the TOPEX/Poseidon and Cassini Educational CD-ROMs, a follow-on NSCAT Educational CD-ROM, a Deep Space Network (DSN) multimedia system, and the Consortium for the Application of Space Data to Education (CASDE) web site. The DDL has also created web based multimedia sites such as the web version of "Welcome to the Planets" (http://pds.jpl.nasa.gov/planets/), and the interactive "Catalog of Government Scientific CD-ROMs" (http://stargate.jpl.nasa.gov:1080/), which allows users to browse the contents of a 500 disk CD ROM jukebox. The DDL also conducts ongoing research of new multimedia technologies, standards and data types, including browse, production and archive formats.

This paper will describe the internal architecture used for capturing and building these multimedia products and the WebCat cataloging system adapted from the Data Archive and Retrieval Enhancement (DARE) software under development for the Defense Special Weapons Agency (DSWA). The cataloging system is based on a flexible model which allows smooth translation into a searchable WWW interface. The interface allows full-text and structured-field database searches, browsing of thumbnail versions of high resolution images, and keyword/value searches through metadata. The HTML interface and database query results are automatically created by Perl scripts from pre-processed browse data. The system supports multi-platform formats and networked access to the data.

2.0 MULTIMEDI A APPLICATION DEVELOPM ENT

The DDL's CD-ROM products are built on an ad hoc basis, tailored to the specific needs of each product's sponsor. However, the DDL has created a hardware architecture, a core development team structure, standards for multimedia data formats and procedures that support concurrent development of several CD-ROM titles as well as web based systems. DDL CD-ROM developers also leverage previous code and technology for use in new titles. In addition, CD-ROM development teams within the DDL are increasingly using the web to facilitate group development where team members are housed at different locations.

2.1 INTRODUCTION

The DDL created its first production multimedia CD ROM, "Welcome to the Planets" in 1993. Several prototype multimedia projects pre-dated "Welcome to the Planets", including the "Planetary Data System (PDS) Home Companion"; a Hypercard application for accessing what was previously paper-based information, and the "Information Systems Newsletter"; an electronic version of the NASA Office of Space Science "Science Information Systems Newsletter." While these early efforts did not make it past the technology demonstration phase, they did lay the groundwork for "Welcome to the Planets" and the products which follow today. Early research into SGML and compound document formats led to pioneering work by the DDL in HTML. Similarly, the DDL conducted research into optical character recognition (OCR) technology and CD-ROM publishing tools, which led to successful products. Early experimentation with multimedia authoring tools such as Hypercard and Supercard led to the eventual acceptance of Macromedia Director as the development tool of choice. As DDL multimedia developers ran up against limitations imposed by Director, they overcame these hurdles by developing custom XObjects/DLLs ("plug ins" for Director) and using new tools such as S(iptX and Java. An ongoing research thrust of the DDL is evaluation of new multimedia authoring tools.

2.2 DEVELOPMENT STAFF STRUCTURE

Following the success of "Welcome to the Planets", the DDL modified its staffing structure to better handle development of several CD ROMs at once. Whereas "Welcome to the Planets" was developed at any given time by one software engineer, follow-on titles employed small teams where specific responsibilities were delegated to specialists. With the recent development of the NSCAT Educational CD-ROM, this team philosophy has been further refined to its current structure

CD-ROM products are currently handled by a development team consisting of:

- Task Manager - Responsible for developing and maintaining schedules, managing team members, allocating task resources, and coordinating with external organizations.
- Lead developer -- Responsible for overall disk design, programming, art direction, integration of content, and coordination of other team members.
- Artist/animator -- Responsible for development of original content (animations and illustrations), as well as assisting the lead developer in the design of the user interface.
- Researcher - Responsible for research and acquisition of existing content (images and video), and negotiating copyright permission agreements with content providers.

In addition, the DDL uses subcontracts with both in-house and external individuals and organizations to fulfill specialized roles on a need basis. An example of a specialized role is narration of CD ROM animations, which is fulfilled by an outside voice artist.

The DDL has also hired staff specialists who can assist development of all CD ROM and web based multimedia products. These skills include CD-ROM premastering and production issues, Windows programming, Perl programming, digital video editing, and UNIX system administration.

2.3 DEVELOPMENT HARDWARE

The DDL has developed a hardware architecture for multimedia production and delivery based on three basic types of systems:

1. Servers to serve up the indexed and processed multimedia data on the web and to control large storage devices.
2. Production workstations for multimedia content creation and CD recording.
3. Client workstations for day-to day use by the staff and for product testing.

The current DDL servers are all hosted on UNIX workstations. Experience has shown that it is mandatory to isolate large storage devices on separate workstations as shown in Table 1. Experience with jukebox systems, both CD jukeboxes and magneto optic jukeboxes has been disappointing. The DDL's magneto optic jukebox is prone to losing its memory of the contents of the platters. A recent power outage over the west coast completely destroyed the catalog of disk contents. Additionally the DDL has been unable to successfully set up a caching system. As a result, the unit spends a great deal of time thrashing (mounting and dismounting platters). The DDI.'s 500 disk CD ROM jukebox has been plagued with sporadic robotic and reader failures which are difficult to isolate and repair.

Table 1 - DDI Servers

Server Name	Server description
Stargate	Sparc 10 general purpose server with 12 Gbyte hard disk and CD premastering system.
Starbase	Sparc 4 with Pioneer DRM-5004X 500 CD jukebox.
Starhawk	Sun workstation with Kubik 240 CD jukebox.
Starlet	Sparc 5 with HP 88 Gbyte Magneto Optic Jukebox.

DDI, production workstations (see Table 2) are tailored 11101110 their function in the spectrum of multimedia development processes and therefore span a range of operating S)'sfl'ills, For example, most animation, illustration, and authoring is done on Macintosh workstations, owing to the rich graphic arts and content creation tools available on the Mac. However, the DDI, also employs Windows 95 and NT workstations for MPEG encoding, optical character recognition, some authoring, and Java development, owing to the advantage of Windows tools in those areas. A Silicon Graphics Indigo 2 was recently purchased to host Wavefront software for development of high-end scientific animations and visualizations. Specialized input devices such as video digitizing boards and scanners are also widely used. This variety of different workstation types and peripherals might portend a networking and file interchangability nightmare, but careful use of standardized multimedia data formats (discussed in the next section) and intelligent design of the supporting network has minimized problems. While intelligent planning of the DDI hardware architecture avoids most problems, the DDI, is also quick to employ promising new technologies, such as the Iomega Jaz removable drives, to solve near-term problems. CD Recordable (CD-R) discs are used for several purposes including archive, small-volume data distribution, and testing. The 1)1)1. is in the process of designing a capability for unattended recording of up 1(140 CDRs during off-shift hours. This will require either a transporter mechanism and high speed recorder or jukebox with recorder installed.

Table 2 - DDI, Production Workstations

Workstation Name	Workstation Description
ScannerPC	Pentium 90 with Adobe Capture and Panasonic duplex scanner.
MpegPC	Pentium 90 with Media Lab Suite MPEG encoder and CD-recorder.
Audiovidium	Macintosh Quadra 950 with VideoVision capture board and JVC CD recorder.
MediaMac	PowerMac 9500 with Media 100qx video capture board.
ScanMac	PowerMac 7100 with HP color scanner.
Megabyte	SGI Indigo 2 with Wavefront software

2.4 MULTIMEDIA DATA FORMATS

The key components of 1)1)1. multimedia products include images, video and animation clips, audio files and compound documents.

The adoption of the Compuserve Graphical Interchange Format (GIF) by the developers of Mosaic and further standardization on the JPEG and GIF image formats by the Netscape and Internet Explorer browsers has finally brought an element of standardization to image formats. However most products are not generated in either of these formats, but are produced in TIFF or some other binary format. The DDI maintains an archival version of the original product with a descriptive label indicating the file format. The archival version is processed with graphics processing tools such as Image Alchemy to create a heirarchy of image sizes for different viewing purposes in multimedia presentations. DDI experience has shown that an **icon** of about 36 x 36 pixels is too small to be a useful distinguishing representation. Instead the 1)1)1, provides a **thumbnail**, which is on the order of 100 x 100 pixels in GIF format, a **browse** image which is screen-sized, 640 x 480 pixels in JPEG format and a **full-resolution browse** image in JPEG format. Some productions, such as the TOPEX/Poseidon Informational CD-ROM visual table of contents, utilize a **small-browse** image of 320 x 240 pixels for rapid browsing through image collections. Image files which exceed about a megabyte are tiled into smaller pieces for display and downloading.

The DDI has standardized on the QuickTime digital video format for both video and audio data. It is the only format currently in wide use that works seamlessly across Windows, Macintosh and Unix platforms. QuickTime players for Windows and Mac platforms have been licensed by the DDL from Apple Computer and are distributed on CD-ROM products for free. The QuickTime movies are created from a variety of video editing, 3D animation, and video effects applications on both Windows and Macintosh computers. The DDL internal standard is to capture at a frame size of 320 x 240 pixels, 30 frames per second, 24 bit color and 22-kHz, 8 bit monaural audio. Output products use the same frame size (i.e. 320 x 240 pixels), but are reduced to 15 frames per second and use the Cinepak compression codec with key frames every second. The movie data rate is limited to 220 kilobytes per second, roughly the throughput of the average double-speed CD ROM reader. Although the maximum theoretical throughput of a double-speed CD ROM reader is 300 kilobytes per second, testing and user feedback indicated 220 kilobytes per second is a more reasonable "real-world" value. The DDI has also experimented with digital video browse files at various rates and has found that ? frames per second produces a good browse rendition of the original video for most products. Table 3 provides file sizes for a variety of QuickTime clips.

Table 3 - QuickTime movie sizes

Movie Name	Data Size	No. of frames	Elapsed time	Avg playback rate	Content
Climate.mov	14.5 MB	1278	1:25 min	"173 KB/sec	anim+video
Launch.mov	17.0 MB	1533	1:42 min	171 KB/sec	anim+video
Surface.mov	5.3 MB	918	1:01 min	90 KB/sec	anim
Systems.mov	10.9 MB	809	0:54 min	206 KB/sec	video

The recent release of Microsoft Active Movie with support for MPEG-1 playback in software at CD-ROM data rates provides hope for a higher level of video presentation in future products. The quality of MPEG-1 scientific multimedia presentations is excellent, retaining approximately the information content of NTSC video. The data storage requirements are about two thirds those of QuickTime Cinepak movies, averaging a little under 10 megabytes per minute at a data playback rate of 150 kilobytes per second.

The new QuickTime VR and VRML technologies provide exciting potential for interactive manipulation of scientific and informational data products. The DDL uses QuickTimeVR object-view for presentation of spacecraft models, and 3D views of geographical areas (e.g. the surface of Mars, Venus, Pasadena, Omaha). Object-views can be very large (10 to 20 megabytes depending on image size and compression) because of the large number of frames required for smooth viewing (648 images for a full 360 x 180 degree view). The QuickTimeVR panorama-view is used for presenting panoramic land views, such as the surface of Mars from the Viking Lander or historical sites such as the Mormon Cemetery in Omaha, Neb. These files are reasonably small, less than a megabyte for most of our subjects. The new panorama and object making software which Apple is providing should bring this technology to orders of magnitude more users.

Audio narrations for animations and videos are captured using Macromedia SoundEdit. The DDL's early products used monaural audio captured at 22 kHz with 8-bit sampling. The resulting products exhibited a substantial amount of noise. The DDI is in the process of defining new procedures that will use 16 bit sampling converted to 8-bit using Deck II software from Macromedia or other sound editing software. The audio files are captured and stored in AIFF format for compatibility with Macromind Director. For wider delivery on the web, the AIFF files are converted to audio only QuickTime movies.

The DDL produced a conference CD-ROM for the International Geoscience and Remote Sensing Symposium (IGARSS) in 1994. Two technologies then just emerging were chosen to present the conference abstracts, papers and digital data submissions. The abstracts were scanned on an OCR scanner and put onto a conference home-page for on-line access and retrieval using WebCat (see Section 3). Several hundred conference papers were submitted by their authors in postscript format and were *distilled* to Adobe Acrobat Portable Document Format (PDF). Papers not submitted electronically were scanned and converted to Acrobat format using Image Alchemy. The resulting Acrobat files were distributed on CD-ROM and were available for downloading from the conference home page. Acrobat readers for several platforms were distributed on the CD-ROM. More recently the DDL has used Acrobat's PDF format for presenting curriculum on educational CD-ROMs and for producing rich compound documents with high-resolution color and monochrome imagery. Adobe Capture is being used to digitize NASA publications and mission documentation for special CD-ROM projects. The DDL is also evaluating Adobe Amber technology for presenting large PDF files over the Internet.

However, the use of Adobe PDF format has not been without problems. Users of the NSCAT Educational CD-ROM continue to report problems installing Adobe's Acrobat Reader application on the Macintosh, and PDF files are quite large compared to competing portable document formats. The DDL invested significant resources trying to solve the Acobat Reader installation problem, but a lack of cooperation from Adobe and licensing limitations have prevented a satisfactory solution. As a result, the DDL is evaluating other portable document formats and believes the *stylesheet* extensions proposed for the HTML 3.2 standard hold significant promise.

2.5 DEVELOPMENT PROCEDURES

The DDL has established common procedures for the development of its CD-ROM products. A standardized JPL software development lifecycle is followed, in addition to special procedures for sub tasks such as video digitization. The development of these special procedures serves two purposes: to reduce development time and cost by streamlining repetitive processes, and to allow different DDL staff personnel to quickly learn new skills.

In its role as a research and development laboratory, the DDL conducts ongoing research into multimedia technologies and makes an active effort to integrate these new technologies into its products. What often results are procedures to be followed by all developers. For example, the DDL has unofficial guidelines for CD-ROM specific issues such as digital video compression settings. The DDL also actively supports continuing education of its members, including regular meetings, journal reviews, coordinated attendance of classes and conferences, and proliferation of knowledge through electronic mailing lists, seminars and publication of journal articles.

2.6 TECHNOLOGY

The overall CD ROM design philosophy is to leverage as much existing code and technology from previous CD-ROMs as possible, discard weak or unused features, and constantly improve. All of this is done within the confines of the specific CD-ROM title requirements, which are established with the input of the customer (*i.e.*, the NASA project sponsoring the CD-ROM), and the target user (e.g. middle-school science teachers.)

Most ongoing CD-ROM titles are developed in Macromedia Director, and utilize a similar architecture. This allows the reuse of much Lingo (a Director's internal scripting language) code and external plug-ins (Xtras, XObjects, XCODs, XCMDs and DLLs). A generic engine, originally developed for the NSCAT Educational CD-ROM, has been re-used in the Cassini Educational CD-ROM and the follow-on NSCAT Educational CD-ROM.

The development of the TOPEX/Poseidon Educational CD ROM resulted in pioneering work in the areas of multimedia metadata modeling and "author-once" publishing.[2] *Author-once* implies that infomation is authored once, and can be published many times for different formats such as print, CD-ROM and the web. In this case, an architecture was created whereby the essential multimedia data elements (images, video, sound, and text) and the links between them were maintained in a master content database. Tools were then developed both in C and in Macromedia Director to extract the information from the database and generate various versions of the presentation (see Figure 1.) This flexible architecture allowed easy changes to content. To make a change to the presentation, only the master database and appropriate content file had to be changed. The change was then automatically made to the HTML, Mac CD-ROM and Windows CD-ROM versions of the presentation. In order to allow such flexibility, the developer created metadata descriptions of multimedia data elements, the links between them, and even the layout of the user interface. The metadata description files were written in the Object Description Language (ODL), a metadata description language used by the DDL in its other products and described in more detail in the next section.

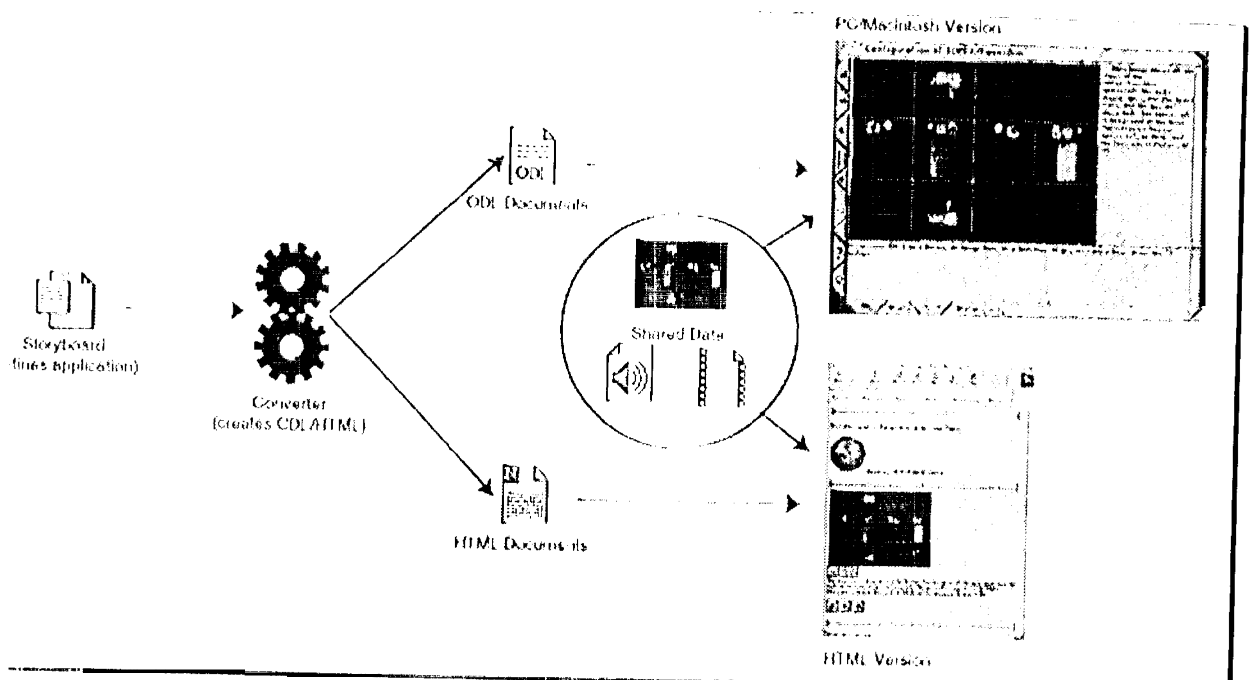

Figure 1. TOPEX/Poseidon Educational CD-ROM Development Architecture

To keep pace with the features and capabilities of commercial CD-ROMs in similar categories, the DDL stays abreast of changes in multimedia technologies and makes an active attempt, whenever possible, to infuse this technology into new CD ROMs. For example, "Welcome to the Planets" and the NSCAT Educational CD-ROMs were CD-ROM products with no links to the Internet. Migration to the web began with a port of "Welcome to the Planets" to the web. The TOPEX/Poseidon Educational CD-ROM followed by offering a web version of the Director application on the CD-ROM. Both the follow-on NSCAT Educational CD-ROM and the Cassini Educational CD-ROM will employ active links to the Internet. This use of "loosely-coupled" hybrid CD-ROMs will allow users to download near-real time data from spacecraft data systems for use in the CD-ROM application.

2.7 USE OF THE WEB TO SUPPORT CD-ROM DEVELOPMENT

The ease of use of the web and the proliferation of groupware has greatly assisted the development of DDL CD-ROMs. Currently, the DDL employs restricted-access web pages for each of its CD-ROM titles. These developers' pages serve as a repository for shared resources and information. This shared information includes storyboards, in-progress illustrations, animations, screenshots of the CD-ROM user interface, task schedules, and links to related pages. In addition, the current working version of the web version of multimedia products is always available to team members.

The DDL has also begun design of a video server to support concurrent development of all CD-ROMs. During the parallel development of the TOPEX/Poseidon Informational and the NSCAT Educational CD-ROMs, it was discovered that much of the video and animation material was usable on both disks. In addition, DDL developers discovered problems archiving working video and animation files. Working files are distinct from release versions of video and animation files. Release versions -- those that appear on the finished CD-ROM -- have been compressed to the data formats described in Secion 3 and are generally much smaller in size. However, these files cannot be edited since the Cinepak codec uses temporal compression. Changes and edits must be made to the **project file**, which consists of working files. The working files are typically uncompressed raw video and audio files, and are significantly larger in size than the release file. As a result, requirements were developed for a video server which could solve both the archive and multiple-use problems. The DDL is currently finalizing the design of the video server system. It is hoped that when completed, this video server system, or a portion of it, will be made available to the public for accessing NASA video and animation archives. Long-term plans are to expand this video server to include other types of mutlimedia data, including sound files, 3D models, and images.

3.0 WWW-MULTIMEDIA DATABASE DEVELOPMENT

3.1 INTRODUCTION

In addition to its multimedia CD-ROM/web-sites, the DDL is actively involved in the development of several web-only multimedia tasks. The principal task in this area is *WebCat*, derived from a system JPL began developing in 199? for the Defense Special Weapons Agency's Data Archival and Retrieval Enhancement (DARE) task. WebCat is a software environment for indexing, cataloging, retrieving, and displaying multimedia data archives over the web. Designed and implemented initially for the DARE task, the adapted WebCat software is also being used by many other JPL organizations for access to web-based information repositories. DARE data consist of technical documents, photographs, technical diagrams, video, numeric data, tabular data, and software. The data set is quite large -- on the order of several terabytes. The principal objective of WebCat/DARE tasks is to archive multimedia data, and provide an on-line data system that allows users to effectively and efficiently access the data. The key to meeting this objective is the design and implementation of a metadata model and catalog that adequately descibes all the archive data contained within it.

3.2 CATALOG AND METADATA MODEL

The catalog system model adopted by the DDL includes five major components as illustrated in Figure 2. The Data Dictionary is a system resource which defines all the metadata parameters in the system. These parameters include definitions of object types stored in the system (image, animation, video. clip) as well as descriptive parameters (date, location, author) for those objects. The Guide (or encyclopedia) is a repository of domain specific knowledge for the. information stored in the. catalog system. For example, one entry in the Data Dictionary might be SPACECRAFT_NAME. Standard values for SPACECRAFT_NAME include MAGELLAN, GALILEO and CASSINI. The Guide will contain extensive descriptions of the MAGELLAN, GALILEO and CASSINI missions. The Directory contains a description of each major category of data object stored in the system. Sample directory entries for a multimedia catalog might include planetary visualizations, weather movies or spacecraft models. The inventories contain entries for specific data object types (image, movie, diagram, document, software, table) and identify individual data objects by url identification, allowing the data to be stored locally or remotely.

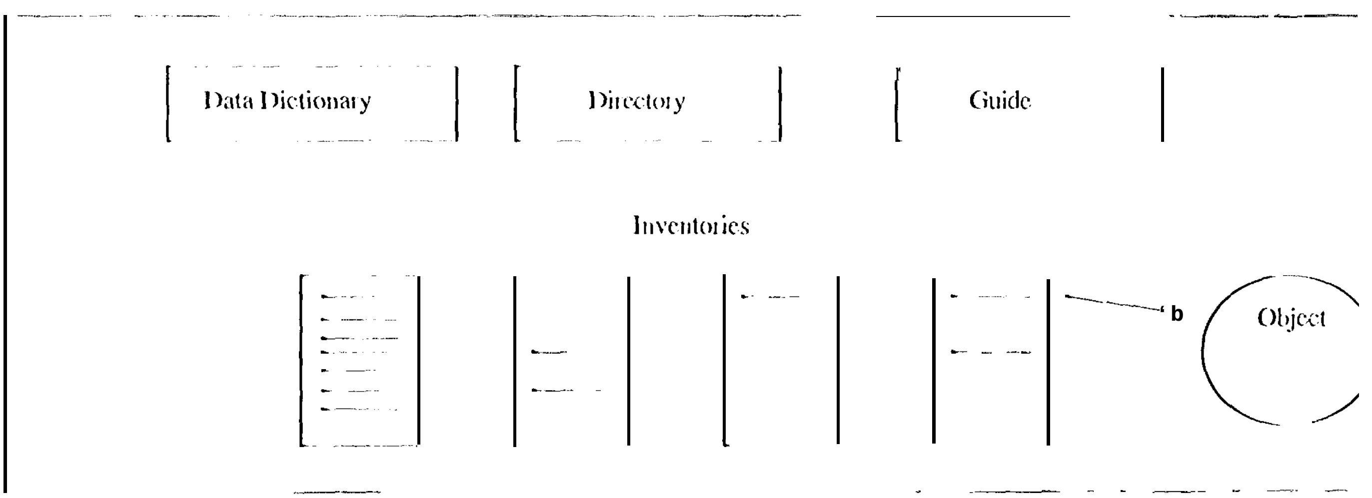

Figure 2 Catalog Model

Entries for all the components of the system model are prepared and stored using the Portable Data Specification (PDS) architecture. This standard was created by the Planetary Data System and extended for multimedia data types by the DARE task The PDS standard utilizes the Object Description Language (ODL) as a language for metadata. The ODL labels are written in ASCII text, which can be easily read by users, modified by simple text editors, and parsed by tools written in Perl or C. The

PDS standard requires a distinct data product label for each individual data product in order to describe the organization, format, content, and quality of each data product.[3]

With this model, the architecture offers versatility and expandability to interface with any type of database or web indexing application.[1] However, the creation of metadata for the data objects is not without problems. In order for the rmd-user system to allow simple, effective, and thorough searches, the metadata must be carefully designed. Keywords and objects must be thoroughly reviewed to ensure they properly describe their data, This is a time consuming process. For example, defining a data type such as a document or movie file may take a month or more of discussion to identify each important keyword and its associated values.

3.3 WEBCAT SOFTWARE SYSTEM

The WebCat software builds HyperText Markup Language (HTML) pages of results from searching indexes of metadata. These HTML pages produced both at setup and runtime are what makes the system unique. All the HTML pages are generated by custom Perl scripts while the underlying WAIS-sf search engine generates and utilizes the indexes. The custom software is composed of four main subsystems:

1. Configure - Prepares the WebCat software environment using configuration files to identify data inventories and supporting information.
2. Setup and Setup Sequencer - Creates the static keyword/value HTML pages, WAIS indexes, and Perl arrays needed for runtime execution. Setup is run on whole inventories while Setup Sequencer is run when data products are added, updated or deleted.
3. Runtime - Executes and summarizes WAIS searches, displays information about a single Inventory or supporting information itc.ills, and dynamically tags the displayed label.[4]
4. Order system - Collects data which were selected by a user while searching the system and produces a CD-R for a distribution.

3.4 SETUP

The setup module of Webcat is a pre-processing and preparation step. The setup component's entire purpose is to preprocess the meta-information stored in the labels and create the input data structures needed by the runtime & layout component.

[meta-info] -> (setup) -> [data structures] -> (runtime) -> (browser)

Fig 3 - Webcat Overview

Setup actually consists of six components. The components are structured for speed and parallelism. Each component has clearly defined input data structures, dependencies and output data structures. L - Locates all the meta-information label files; AI - adds a unique Webcat specific ID and search-engine specific headline to each meta-data label; T - Creates a tm<}j-table for use in H; H - Creates the GUI interfaces: full text, structured field search, and keyword-browse HTML pages; W - Starts the database full-text and structured field indexing; R - Produces a set of three quick-lookup tables for use by the Layout component. Modules L, AI, and T are for 'internal' use only. The H, W, R modules output data that is used in the Runtime system.

3.5 RUNTIME & LAYOUT

Whereas older methods of creating WWW archives required manual preparation and maintenance of source material, Webcat dynamically generates the WWW interface. The Run Time & Layout module of Webcat has three ways for users to interact with the underlying metadata and eventually view (Layout) the information they're looking for.

GUI	Module	GUI	Module
Full-Text Search	runtime	list of titles	layout
Structured-field Search	runtime	list of titles	layout
Keyword/Title Browse	Output of Setup II	list of titles	layout

Fig 4 - Webcat GUI interfaces

All of these approaches eventually result in a listing of titles. The Runtime module reads all of its input from the output of the Setup components. The two Search GUI's rely on the output of Setup W,]<,]]. The Browse GUI relies on the output of R and H. The Full-text Search GUI is a simple HTML/CGI forms interface to the database search engine (WAIS in this example). Search terms are simply passed directly to the database. The Structured Fields GUI is a more complex HTML/CGI forms interface to the structured fields database search engine. (WAIS-SF in this example). Each index field in the database is listed. If there is a 'manageable' number of terms for that indexed word, they are all listed in a scrolling list, allowing the user to select only terms that are in the database. This proves very effective for fields with standard values (e.g. n SUBJECT keyword that contains only the values FICTION, PERIODICAL, NON-FICTION). After selections have been made in the form, the results are parsed and a possibly complex query is generated and passed to the database.

The Keyword/Title Browse GUI is relies on the output of Setup H. The H component processes all the indexed keywords and produces a listing of hierarchical titles for that keyword.

Setup H: Indexed Keywords	Setup H: Ranges	Setup H: Values	Layout: Values
1. Event Name (204) 2. Subject Type (20)	Event Names 1. Ant - Mob (100)" 2. New - Wiz (100)" 3. Yes - Zig (4)	Event Name Values 1. Yes (1) 2. Zebra (3) 3. Zero (10) 4. Zig (34)	Event Name : Zebra 1. A Zebra and Ant 2. An African Story 3. Safari Tales

Fig 5 - Keyword/Title Browsing

Upon selecting a title from the list, Layout looks up the meta information using the tables created by setup R. This information is then displayed on the user's browser as specified in an inventory specific manner. For example: a hypothetical movie database had it's metadata converted to a simple Webcat format consisting of TITLE, PLOT, RATING, MOVIECLIP and ARCHIVE_LINK. Upon reaching the Layout, the browser could display the TITLE text, PLOT text, a RATING icon, an inline QuickTime MOVIECLIP and a hyperlinked ARCHIVE_LINK to the rec.arts.movies.reviews archive site.

3.6 ORDER SYSTEM

Locating the information of interest is the first objective in developing a useful interface, and Webcat offers that with a variety of easy-to-use search options. The second objective is to make the data available to the user. While electronic file transfer is still used, the massive quantity of information and large data product sizes, have made an Order System for Webcat a requirement. At each phase of a search where a list of titles is displayed, or a layout page of a title is presented, there is an order button that is part of the HTML/CGI interface. The three order options include:

1) Selecting all titles returned from a query
2) Selecting from a list of titles returned from a query
3) Adding a single title from the layout page

When an order is first placed using one of the order options mentioned above, three files are created; an order label, a customer label, and a content file. The order label contains several customer modifiable fields. These include the name of the order, the medium on which the customer wants to receive the order (e. g. CD R, 8mm tape), and the option of ordering both the archive and the browse data or just the browse data. The customer label consists of information about the customer placing the order I, such as name, address, phone number, e-mail address, and the customer's access privileges if it the order is being placed on a server where there is sensitive or restricted data. The content file contains the list of titles to be ordered These three files are linked, and the order is tracked by a unique. order II).

The content of the order, and order information can be viewed and modified at any time This View/Modify Order page also has the total size of the order, and based on the medium selected, the number of media required to fill the order. A restriction on the "reasonableness" of an order is also implemented. The reasonableness criteria is based on system resources, data size and suspected usage. An example of a reasonable order might be one that requires less than five CD Rs to fill.
Once a customer has finished an order, they submit it to be processed. The content file is automatically broken up into volumes based on the medium selected. The operator views the submitted orders through the browser , and based on the hardware configuration, sets the processing date. and time for each submitted order. The amount of operator intervention required to place an order is based on the medium selected, and the hardware configuration. Along with the data ordered by the customer, a validation file is created to ensure the data was accurately copied to the medium. A packing list of the titles ordered is generated in HTML with links to the metadata and associated data files. So, for example, if an order was placed on a CD-R, a customer could view the packing list with a browser, select a link on the page, and go to the data on the CD-R.

For servers with limited or restricted access, the ability to create proxy orders has been implemented. A staff member using the same interface as a customer has the added capability of creating a temporary customer label and placing an order for a customer who does not have access to the server, but has been given permission to order data.

4.(ACKNOWLEDGMENTS

The work described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, 01 otherwise does not constitute or imply its endorsement by the U.S. government or the Jet Propulsion Laboratory.

We would like to thank the following individuals for their support - Ann Bernath, David Bernath, John Gaa, Adrian Godoy, Sue Hess, David Hecox, Steve Hughes, Jason LaPointe, Carol Miller, Randall Oliver, Helen Qian.

5.0 REFERENCES

1. Hyon, Jason J., Borgen, Rosana B., "Data Archival and Retrieval Enhancement (DARE) Metadata Modeling and its User Interface", *The first IEEE Metadata Conference Proceedings*, March 1996.
2. Hecox, David and Martin, Michael, "The TOPEX/Poseidon Informational CD-ROM", *Information Systems Newsletter*, vol. I 1996, Issue 38.
3. 1'1)S Standard Reference, J]'], Internal Document D-7(169, Jet Propulsion Laboratory, November 1992
4. Hughes, J. S, Bernath, A. M., "Planetary Data System Data Model as Applied to WWW interfaces (WEBCAT)", JPL & NASA Case No. NPO 1966-/, October 1995